On Aligning OpenIE Extractions with Knowledge Bases: A Case Study

Kiril Gashteovski$^{1,2}$  Rainer Gemulla$^1$
Bhushan Kotnis$^2$  Sven Hertling$^1$  Christian Meilicke$^1$

$^1$University of Mannheim  $^2$NEC Labs Europe
Germany
Open Information Extraction (OpenIE)

Extract triples from natural language text in unsupervised manner

“Michael Jordan, who played for the Chicago Bulls, was born in Brooklyn.”

(“Michael Jordan”; “was born in”; “Brooklyn”)
(“Michael Jordan”; “played for”; “Chicago Bulls”)
OpenIE

Ambiguous: triples whose elements are strings

("Michael Jordan"; "played for"; "Chicago Bulls")

"Michael Jordan" → 16 entities in Wikipedia
"Chicago Bulls" → 2 entities in Wikipedia
"played for" → string, does not have precise meaning

Knowledge Bases (KBs)

Unambiguous: triples whose elements are unambiguous concepts

(Michael Jordan; dbo:team; Chicago Bulls)

Michael Jordan
Chicago Bulls

KBs suffer from low coverage
OpenIE and KBs

• OpenIE can be used to construct or enhance KBs

**Question:**
How the information in OpenIE triples relates to the information in a KB?

• We *manually evaluate* the semantic relatedness between OpenIE and KB
  • We used the OpenIE corpus **OPIEC** and **DBpedia** KB
OpenIE triples and KB facts with same arguments

• Inspired by the **Distant Supervision Assumption (DSA)**
  • **KB-Hit**: for an OpenIE triple there is a KB fact with same argument pair
  • **DSA**: They express the same information

  \[
  (\text{Michael Jordan}; \ "played for"; \text{Chicago Bulls})
  \\
  (\text{Michael Jordan}; \text{ dbo:team}; \text{Chicago Bulls})
  \]

  “*sports teams or clubs that the subject currently represents or formerly represented*”

• Key assumption: used for bootstrapping OpenIE and expanding KBs
OpenIE triples and KB facts with same arguments

• KB-hit may not always have equivalent semantics as the OpenIE triple
  • for each KB-hit, we differentiate four hit categories
OpenIE triples and KB facts with same arguments

• Split the OpenIE triples in two groups:
  • *Is-a relation*: indicate *types*; e.g. (“Michael Jordan”; “be”; “basketball player”)
  • *All relations*: all other OpenIE triples

• Manually assign *best hit category* on each OpenIE triple
OpenIE triples and KB facts with same arguments

• Most OpenIE triples express the best hit
  • Though, the OpenIE triples tend to be more specific
• OpenIE triples contain more fine-grained type information

(a) All relations

(b) Is-a relation
Expressibility of an OpenIE triple with a DBpedia fact

• To what extent an OpenIE triple contains information relevant for DBpedia?

• Three possible expressibility levels

(Michael Jordan; “played in”; NBA) \(\text{Fully-Expressible}\) (Michael Jordan; dbo:league; NBA)

(Michael Jordan; “played for Bulls in”; NBA) \(\text{Partly-Expressible}\) (Michael Jordan; dbo:league; NBA)

(Michael Jordan; “be fielding a NASCAR team with”; Bubba Wallace) \(\text{Not-Expressible}\) X
Expressibility of an OpenIE triple with KB formulas

\[(\text{Michael Jordan}; \text{“played for Bulls in”}; \text{NBA})\]

\[\Downarrow\text{Fully-Expressible}\]

\[(\text{Michael Jordan}; \text{dbo:league}; \text{NBA}) \land (\text{Michael Jordan}; \text{dbo:team}; \text{Chicago Bulls})\]
Expressibility of an OpenIE triple with DBpedia

• Expert annotator labeled a sample of OpenIE triples
  • Candidate generation strategies: single KB fact and KB formula
Expressibility of an OpenIE triple with DBpedia

- Most OpenIE triples can be expressed with single DBpedia fact
- **KB formulas** significantly increase the expressibility of OpenIE triples
Expressibility of an OpenIE triple with DBpedia

• Most OpenIE information relevant for DBpedia is **not present in** DBpedia
Takeaways

• **Distant Supervision Assumption (DSA) for OpenIE**
  • Mostly satisfied
  • OpenIE triples tend to be more specific

• **Expressibility** of OpenIE triples with DBpedia
  • Most OpenIE triples are **relevant for DBpedia**
  • **KB formulas** significantly increase expressibility
  • Most OpenIE information that is relevant for DBpedia is not present in DBpedia

• **Transferability**: our findings largely transfer over to other OpenIE systems

[https://www.uni-mannheim.de/dws/research/resources/opiec/](https://www.uni-mannheim.de/dws/research/resources/opiec/)